

社会的期待迎合バイアスと投票参加 ——リスト実験による過大推計バイアス軽減の試み——

論

説

善 教 将 大

1. 背景と問題設定

意識調査の実施には様々な「悩みの種」が存在する。標本抽出の際の偏り (sampling bias) はその代表例である。この問題に加えて近年、多くの研究者に注目されているのが社会的期待迎合バイアス (Social Desirability Bias: SDB) である。SDB とは「社会的に是認されている行動を過大報告したり、社会的に是認されていない行動を過小報告すること」(遠藤 2012: 3)、すなわち回答者が体系的な嘘をつく傾向性のことをいう。違法ないし不法行為に対する質問や、人種あるいは性差別への選好など、SDB は回答者が過敏に反応する質問 (sensitive questions) を尋ねる際に発生しやすいとされる。

SDB の例の 1 つとして、政治意識調査における投票参加の過大推計問題をあげることができる。投票参加の過大推計問題とは、投票参加に関する意識調査を実施した際、調査上の投票率の推定値が、実際の値よりも高くなる問題をいう。表 1 は、ここ数年の間の明るい選挙推進協会による世論調査の結果と実際の投票率を比較したものである。いずれの国政選挙においても、推定される投票率が実際の値より 20 ポイントほど高い。調査モードが一貫していないことなどから、⁽¹⁾ 過大推計の程度にはばらつきが存在するが、投票率の推定値が実際の値よりも高くなる傾向は一貫してい

表1 政治意識調査における過大推計バイアスの実際（衆院選・参院選）

		実際の投票率 (%)	投票率の推計値 (%)	実際の投票率との差 (%)	N (人)
衆院選	第44回	67.5	83.0	15.5	1621
	第45回	69.3	87.2	17.9	1798
	第46回	59.3	74.5	15.2	2155
参院選	第21回	58.6	82.7	24.1	1853
	第22回	57.9	80.3	22.4	1899
	第23回	52.6	71.3	18.7	2019

注) 明るい選挙推進協会による意識調査の結果をもとに筆者作成

る。

SDBは、意識調査を用いて実際の政治行動を明らかにする記述的推論の際のミスを生じさせるのみならず、因果的推論のミスを生じさせる原因ともなる。例えば、ある人物（A氏）が選挙で落選し、その原因が実は、A氏の人種に対する有権者の差別意識にあったとする。そうであるにもかかわらず、すべての有権者が「差別意識は抱いていない」と、意識調査をした際、回答するとどうなるだろうか。この場合、独立変数は無分散なので、有権者の差別意識がA氏落選の原因ではないことになる。⁽²⁾このよう

- (1) SDBの規定因としてよく知られているのが調査モードの違いである。調査員による面接聴取法だとSDBが発生しやすく、郵送法や電話調査だと発生しにくい。表1でいえば、第46回衆院選と第23回参院選では意識調査の実施方法が、面接聴取法から郵送調査に切り替えられているため、回収率などが増加するだけではなく、SDBも軽減されている。試論的に筆者が、過大推計の程度を従属変数に、調査モードの違い（面接：1，郵送：0），実際の投票率，有効回収率を独立変数とする回帰分析を行ったところ（N=6），調査モード変数の係数の符号の向きは負であり，かつ，10%水準で統計的に有意であった。
- (2) ただし質的比較分析（qualitative comparative analysis）の文脈などでは，無分散であっても必要条件とみなせばよいということになるので，無分散デザインは決定的な問題とならない（Goetz and Mahoney 2012=2015）。

な推論の誤りを防ぐという意味でも、SDBの軽減は重要である。

もっとも投票参加の過大推計の原因には、いくつかの見解がある。SDBではなく標本抽出の際の体系的誤差ではないかという指摘はその1つである。つまり政治に関心がない人が意識調査に協力しないからデータ上の投票率が高くなると考えられるのである。⁽³⁾しかし、調査拒否者の絶対数が少なかつた1970-80年代の政治意識調査でも過大推計問題は発生している。⁽⁴⁾集計データの分析を通じてこの問題を検討する研究では (Selb and Munzert 2013), 回収率が過大推計バイアスに与える影響に加えて本稿でいうSDBも重要であることが示されている。日本の意識調査を用いてこの点を実証的に分析する松林 (2015) においても、サンプリングバイアスの影響は限定的であることが明らかにされている。⁽⁵⁾政治関心が低い、棄権しがちな人が調査に協力しない可能性は十分あり得る。しかし棄権者の拒否だけではなく、あるいはそれ以上に、SDBの観点からこの問題を検討する必要がある。

本稿ではSDBを軽減する方法の1つであるリスト実験 (list experiment) を用いて、投票参加の過大推計問題の原因を分析する。早稲田大

ただしそうであっても、この例の場合は、誤った因果推論を行うことになる。

(3) 山田 (2009) はその可能性を指摘している。

(4) 1976年に調査が実施されたJABISSでは、投票率の推計値が93.2%である (DKを除くと約90.9%)。実際の第34回衆院選の投票率は73.45%なので、20ポイント近く過大推計していることになる。この調査の有効回収率は約72%であるので、調査拒否者の絶対数は多くない。なお、JABISS調査の詳細はICPSRのHPで確認できる (URL: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4682> 2015年8月25日アクセス)。

(5) ただし松林 (2015) は、SDBの影響もそれほど大きくはないのではないか、という疑義を暫定的な分析結果を基に呈しており、この点の本稿の主張とは異なる。

学による Computer Assisted Self-Administered Interview (CASI) 方式による SDB の研究など (田中・日野編 2013), 調査モードの比較分析については日本でも実証研究が蓄積されている。例えば西澤・栗山 (2010) は, CASI と従来型の面接聴取法 (PAPI) との比較を通じて, 投票参加経験率を申告する際, 過大に参加経験を報告する SDB が発生することを実証している。しかし, 面接調査法ではないオンライン調査や郵送調査でも, 投票参加の過大推計問題は発生する。さらに, CASI の実施には多額の資金が必要だという問題もある。CASI は個人レベルで実施できる方法ではない。

リスト実験は調査モードの変更ではなく, サーベイ実験によって SDB の軽減を図る方法である。SDB を軽減するサーベイ実験としては, コンジョイント分析やクロススイズモデルによる分析などもあるが, リスト実験はその有用性が実証されていることもあり, 多くの研究者に注目されている⁽⁶⁾。加えて投票参加と SDB に関する実験的手法を用いた先行研究 (Holbrook and Krosnick 2010a, 2010b; Persson and Solevid 2014) の一部でも, リスト実験は用いられている。本稿でもこれら先行研究と同じく, リスト実験によって投票参加の過大推計問題について検討する。

ただし本稿は, 投票「参加」だけではなく投票「方向」も加味した分析を行う点で, これまでの研究より発展的である。有権者が特定の投票先を表明しにくい⁽⁷⁾ということは十分に考えられるが, この点を考慮した分析は

(6) リスト実験に関する妥当性を検証したものとしては, 土屋・平井・小野 (2007) がある。また, 比較的近年のレビュー論文としては Wolter and Laier (2014) がある。いくつかの問題が内在する方法ではあるが, SDB の軽減という点で有効な方法であることは, 多くの実証研究から示されている。

(7) 公明党や共産党など, 多くの有権者の拒否意識が高い政党に投票した場合などがその例として考えられる。

設計が困難ということもあり、ほとんど行われていない⁽⁸⁾。本稿は、投票方向が単純な二項選択型選挙である、2015年5月に投開票が行われた都構想住民投票での投票行動を対象とすることで、方向性も加味した分析を行う。そうすることで、投票行動に係るSDBが参加することに基礎づけられるのか、あるいはそれだけではないのかを判断することが可能となる。

2. 方法論の概略

2.1 サーベイ実験

前節で述べたように、SDBを軽減させる方法としてはいくつかの方法が既に考案されている。本稿が用いるのは、その1つであるリスト実験である。この方法について説明する前に、意識調査と実験を組み合わせるサーベイ実験 (survey experiment) について説明しておきたい。

近年、意識調査に実験を組み込み、因果推論を行う研究が急増している。実験というと一般的には実験室実験 (laboratory experiment) を想像しやすい。しかし本稿のように、意識調査の回答者をいくつかの異なるグループに配分し、それぞれのグループに対して異なる刺激等を与える、といった形で仮説検証を行う研究もある。そのような研究も、実験として位置づけられる。

サーベイ実験にはいくつかの利点がある。第1は対象者の代表性を実験室実験以上に担保できることである。実験室実験は学生サンプルなど、対象者に偏りが存在すると考えられることが多く、知見の外的妥当性に疑義を抱かれることがある⁽⁹⁾。これに対してサーベイ実験は、一般の人で、しか

(8) 設計上の問題もさることながら、特定の候補者や政党に投票することにSDBがかからないという前提があるようにも思われる。しかし、注7で示した例など、投票方向にもSDBがかかる場合はあるだろう。

(9) しかし近年においては、社会心理学などで Amazon Mechanical Turk

も大量の対象者に実験を行うことが通常である。第2に、サーベイ実験は無作為配分を確実に行うことができる。これは、人為的に無作為配分するわけではない自然実験と比較した場合（Dunning 2012）、利点だといえる。

もちろんサーベイ実験に問題がないわけではない。サーベイ実験は、あくまで意識調査でできる範囲内ではしか実施できない。言い換えると複雑な操作を必要とする実験は、基本的には実施が不可能である。また、自然実験ほど知見の外的妥当性が担保されていないという問題もある。特に仮想状況を設定するようなサーベイ実験は、その点を批判されやすいように思われる。

このような問題を抱えつつも、サーベイ実験は有用な知見を提供するツールとなり得ると多くの研究者にみなされている。それゆえに近年、サーベイ実験を用いた分析が政治学では急増している⁽¹⁰⁾。

2.2 リスト実験

リスト実験は、具体的には図1に示す手続きで行われる。第1に回答者を、複数の群（基本的には2群）に配分する。この時、それぞれのグループサイズが十分な大きさであり、かつ、配分方法が無作為であれば、これらは近似する特徴をもつ群となる。例えばA群の男性割合が60%である場合、B群の男性割合も60%に近い値になる。第2に、それぞれの群に対して、異なる項目が列挙されたリストを提示する。一般的には統制群には3ないし4項目のリストを提示する。処置群には、統制群にセンシティブ

を代表とするクラウドソーシングサービスを利用する研究が急増しており、この問題は改善されつつあるように思われる。

(10) 実験の手法を用いた研究は、政治学では計量分析による因果効果の推定の限界が認識されると同時に普及したという経緯がある。その意味で、現在、計量分析が増加している行政学や政策学でも、実験に基づく実証研究が、今後増加する余地は大いにあると考えられる。

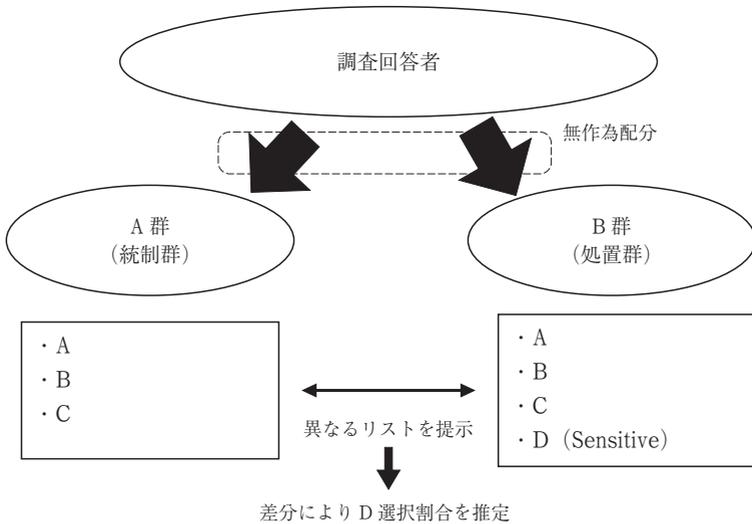


図1 リスト実験の基本モデル

項目を1つ加えたリストを提示する（4ないし5項目）。

第2に回答者は「2」や「3」というように、提示されたりストの中で経験したことがある項目や、あてはまる項目の数のみ回答する。どの項目を経験したのかが調査主体にはわからないようにする。そうすることで回答者が安心してセンシティブ項目を選択できるようにする。

第3に統制群と処置群の差分を推定する。適切な無作為配分が行われていれば、センシティブ項目（D）を除く項目選択数は、統制群と処置群で近似する。例えば統制群の平均項目選択数が1であった場合、同じリストを見せたなら、処置群の平均項目選択数も1となる。そのため、処置群と統制群の差分は、センシティブ項目の選択比率となる。例えば統制群の平均項目選択数が1、処置群の平均項目選択数が1.5の場合、およそ50%の人がセンシティブ項目を選択していることになる。

リスト実験の発想は以上に見たように非常に単純だが、他方で異なるリ

リストを2つ作成し提示すればよいというわけではなく、リストの作り方には細心の注意を払わなければならない。統制群に提示する項目は、いずれもセンシティブではないものにしなければならない。また、選択割合の分散が小さい項目の方がよいとされる (Krumpal et al. 2014)。また、回答者が嘘をつかないという前提が必要だとも指摘されている (Blair and Imai 2012)。これらの問題がある場合、リスト実験による推定は妥当性を失う。

他のSDBを軽減する方法との関連について述べておこう。SDBを軽減する方法は先に述べたようにいくつか存在するが、クロスワイズモデルとの比較から (Krumpal et al. 2014)、また多くのRandomized Response Technique (RRT) との比較から (Coutts and Jann 2011)、リスト実験の有効性は示されている。リスト実験を用いて票の売買といった実態解明が困難なテーマに取り組む研究も、近年見られるようになってきた (Imai, Park and Greene 2015; Gonzalez-Ocantos et al. 2012)。その一方で、リスト実験が他の推定方法よりも劣位になる場合もあるなど (Rosenfeld, Imai and Shapiro forthcoming)、万能道具というわけではないことも同時に明らかになっている。しかし設計に問題がなければ、リスト実験はSDBを軽減させる有効な方法である点については、同意が得られていると言ってよい。

3. データと実験デザイン

3.1 データ

本稿では、2015年5月17日に投開票が行われた「大阪市における特別区の設置についての投票」、いわゆる都構想住民投票後に筆者が実施した意識調査を用いて、リスト実験によって投票参加に係るSDBがどの程度軽減されるのかを分析する。

意識調査は2015年7月1日から7日にかけて実施した。調査票はすべて8(722) 法と政治 66巻4号 (2016年2月)

以下に、選挙や政治とのかかわり合いについての意見が示されています
それぞれの意見について、あなたはどのようにお考えでしょうか

	そう思う	ややそう思う	どちらともいえない	あまりそう思わない	そう思わない	わからない
政府のすることに対して、それを左右する力はない	○	○	○	○	○	○
私は政府に対して何の発言権も持っていません	○	○	○	○	○	○
大勢が投票するのだから、自分ひとりくらい投票しなくてもよい	○	○	○	○	○	○
この項目は左から2番目の選択肢を選んでください	○	○	○	○	○	○

図2 Saticeficer を選別するための質問

て Qualtrics を用いて作成した。回答者の無作為配分も Qualtrics のシステムを利用して行った。調査対象者は、(株)楽天リサーチのモニタに登録している、大阪市在住の20歳から79歳までの男女である。調査依頼は、年齢(20代、30代、40代、50代、60代以上)と性別(男性、女性)の割合が国勢調査の値と一致するように調整した。ただし「20代×男性」と「60代以上×女性」に関しては、十分なサンプルサイズを確保できなかった。そのためウェイトを作成し、それを用いて結果を補正する。総回答者数は Drop サンプルを含めると2387となるが、後述する Saticeficer など調査から除外した回答者が一定程度存在するため、有効サンプルは1804である。

ここで Saticeficer について簡単に説明しておきたい。本稿では、推定結果の妥当性を高める工夫の1つとして、Saticeficer をデータから除外している。近年、オンライン調査に付随する問題の1つである Saticeficer についての研究が進められている(三浦・小林 2015)。Saticeficer とは自身の「効用」を最大化する回答者であり、実際の傾向としては、極値・中点や DK など特定の回答を選択する場合が多い。リスト実験は回答者が正

(11) 調査を開始しなかったサンプルまで含めると2425である。

確にこたえない場合、誤った推定結果を導出することになるので、正確な回答をしない Saticeficer については対象者から除外した方がよいと判断した。筆者の調査では図 2 に示すマトリクス型の質問の中に「この項目は左から 2 番目の選択肢を選んでください」との項目を設け、左から 2 番目以外の回答を選択した人を Saticeficer とし、これを対象から除外した。

3.2 従来型の質問

本稿では、従来型の直接尋ねる形式と、リスト実験による結果の両者を比較することで、SDB がどの程度軽減されるのかを明らかにする。ここでの従来型の質問とは、投票行動を、次のような質問文で尋ねる方法である。

<質問文 1：投票参加>

5月17日に行われた住民投票についておうかがいします。あなたは、住民投票で投票しましたか、それとも棄権しましたか。

選択肢はわからない (DK) とこたえない (NA) を除き、1：行かなかった、2：行こうと思っていたが行けなかった、3：期日前・不在者投票した、4：投票日に投票した、である。実際の分析の際は、1と2、および3と4を統合する。DKとNAは欠損値として扱う。

このように投票参加について尋ねた後に、投票方向、つまりどの候補者ないし政党に投票したのかを尋ねることが政治意識調査では一般的である。ここでの意識調査では、上述した質問で3あるいは4と回答した人に対して、さらに次の質問で投票方向を尋ねている。

(12) 項目の順序はランダムイズしているため、常に図2のように4番目に示されるわけではない。

<質問文2：投票方向>

住民投票で投票した方におうかがいします。あなたは、賛成と反対のどちらに投票しましたか。

選択肢はDKとNAを除き、1：賛成、2：反対、3：その他、である。その他を加えているのは、住民投票が記号式ではなく自書式だったからである。その他、DK、NAの順序は固定し、賛成と反対はランダム化した。その他と回答した人は、欠損とした。

なお、意識調査実施のタイミングが投開票日から約1ヶ月後であることを考慮し、投票参加、投票方向ともにDK選択者に関して、一度だけ、同じ質問を尋ねている（probing）。しかしDK率は低く、かつ再度尋ねてもDKと回答することがほとんどであったためこの措置による分布の変動はほとんどない。有効な回答が得られた場合は、欠損とせず有効サンプルに含めている。

3.3 リスト実験のデザイン

従来質問に加えて、リスト実験による投票参加率の推定も行う。リスト実験は、図2で示すように2群にわけて行うことが多いが、この調査では3群にわける。投票方向も加味した分析を行うためである。具体的には3つのグループに対して、共通の質問文で、提示されたリストの中から行ったことのある項目の数を記入してもらう。質問文は「住民投票の選挙期間中にあなたが行ったこととしては、どのようなものがありますか。以下のリストの中から、行った項目の数のみご回答ください。何を経験したのかは、いっさいご記入頂かなくて結構です。」である。

統制群には、以下に示す4つの項目からなるリストを提示する。これら4つの項目のうち「フラッシュモブへの参加」や「街頭での訴え」は、一

般の有権者はほとんど経験しない項目であろう。投票以外の政治参加には過小評価のSDBが発生するという知見があるが（西澤・栗山 2010），この2項目については一般の人が経験しないことが容易に仮定できる⁽¹³⁾。残る2つは一般の人が比較的良好に経験する項目である。「家族と話すこと」や「ニュースに目を通すこと」は日常的な行為なのでセンシティブ項目ではないと考える。

- ・投票啓発のためのフラッシュモブに参加しパフォーマンスした
- ・特別区や大阪市の今後について、家族や友人と話し合った
- ・都構想に関するニュースや記事に目を通した
- ・自らが街頭に立ち都構想のメリットやデメリットを訴えた

続いて、第1の処置群には上述した4つの項目に、センシティブ項目として考えられる「反対への投票」項目を追加したリストを提示する。また第2の処置群には「賛成への投票」項目を追加したリストを提示する⁽¹⁴⁾。センシティブ項目は、統制群に提示した4つの項目の末尾（上から5番目）に追加した。

(13) 街頭に立って演説していたのは主に政治家であること、またフラッシュモブに参加するには高度なダンス技能が必須だからである。明らかに参加のコストが高い（あるいは実質的には一般の人は参加できない）政治参加形態なので、SDBの問題をそもそも考慮する必要がない。

(14) 反対あるいは賛成票を投じたとする際、選挙で、という表記としているが、厳密には「住民投票」である。もっとも、質問文が「住民投票の選挙期間中に」となっており、かつ、賛成あるいは反対票を、としているので、ここでいう選挙が住民投票を意味することは十分に想定できる。実際に、都構想選挙というワードを用いて住民投票を語る人は多い。Twitterのハッシュタグの中にも「#大阪都構想選挙」がある。

<処置群 1 >

- ・「反対だ」という意思を反映させるために、選挙で反対票を投じた

<処置群 2 >

- ・「賛成だ」という意思を反映させるために、選挙で賛成票を投じた

3.4 実験の手順

実験は、図3に示すフローチャートのもとで行った。まず回答者全員に、性別や年齢、居住する区といった一般的な属性に関して尋ねた。その後、投票参加に関して、従来型の質問を用いて尋ねた。リスト実験はその後に実施した。

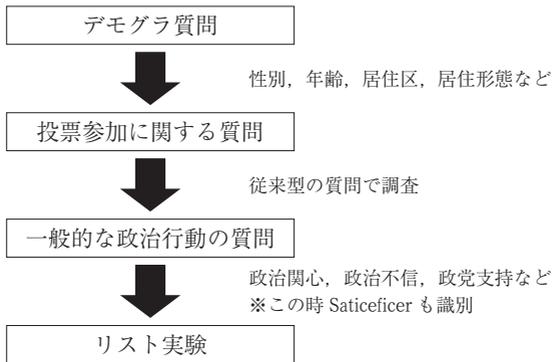


図3 リスト実験の手順

直接尋ねる質問による参加率の推定とリスト実験による推定の比較を、本稿のように同一サンプル内で行うことは一般的ではない。通常は、回答者を複数のグループに無作為に分けた上で、あるグループには直接的に尋ね、他のグループにはリスト実験を、という形で行われる。同一サンプルに尋ねた場合、前の質問がリスト実験の結果の影響を受ける可能性を完全に除去できないためである。

この問題があるにもかかわらず、本稿が同一サンプルに調査を実施した理由は、第1に、4つのグループに分割すると、1グループあたりのサンプルサイズが小さくなってしま⁽¹⁵⁾う。第2に、従来型の質問とリスト実験の間に設問を介在させることで、質問間の影響はある程度軽減される。第3に同一サンプルに尋ねることで、SDBの存在を、より直接的に示すことが可能となる。同一サンプルに尋ねることの利点を、本稿は欠点以上に重視した。

加えて、従来型の設問を事前に尋ねることが、必ずしも歪みをもたらすわけではないことも指摘しておきたい。いわゆる事前の質問が後の質問に与えるキャリアオーバー効果 (carry over effect) に関する研究は、連続する質問を用いることが多い。連続しなければ、強いキャリアオーバー効果は発生しないように思われる。また緩衝項目を挿入することでキャリアオーバー効果が緩和される場合もある。どの程度の数が必要かは場合によるが、緩衝効果を検討する研究では8問程度が挿入されており (Ottati et al. 1989), その結果、質問間の独立性が担保されることが実証されている。このように緩衝項目を挿入することで、事前の質問の影響は、ある程度緩和できると考えられる。

ただし、緩衝項目の挿入という方法には批判もある。山口 (2005) では、緩衝項目挿入の効果そのものについての見解が定まっていないこと、キャリアオーバー効果は必ずしも是正されるべき問題ではないこと、⁽¹⁶⁾ 調査票のまとまりがなくなることなどが指摘されている。これらが示唆する通

(15) リスト実験を行う際、1グループあたりのサンプルサイズを、最低でも400以上、確保すべきであることは Krumpal et al. (2014) などで指摘されている。

(16) ただしキャリアオーバー効果を許容することで誤差項が独立ではなくなるケースが出てくる場合は、キャリアオーバー効果は抑止すべき問題だということになる。

り、常に緩衝項目の挿入が推奨されるわけではないが、本稿はSDBの軽減を目的としており、キャリアオーバー効果は抑制しなければならない。そのためには質問間の関係を独立したものとして回答者に認識させる必要がある。緩衝項目の挿入は、そのための手法の1つとして活用できる。本稿では、10問程度の質問を緩衝項目として挿入することで、キャリアオーバー効果を抑制する。

4 実証分析

4.1 参加バイアスの検証

投票行動に係るSDBはリスト実験によってどの程度軽減されるのか。まずは本稿で用いる調査結果にもSDBが存在することを確認しよう。表2を見ればわかるように、直接尋ねた場合「投票した」人の割合は約81.4%となる。実際の住民投票の投票率は66.8%なので14.6ポイントほど過大推計していることになる。14.6ポイントという値の大小を論じることは難しいが、表1の結果と比較するとやや小さい。その理由は定かではないが、オンラインモニタ独自の傾向性や住民投票での投票率の高さなどが考えられる。投票率が高いとSDBも必然的に小さくなるのであろう。

次に従来型の質問の参加率推定値とリスト実験のそれを比較する。従来型の質問の推定値は既に述べた。リスト実験の推定結果は反対が約22.6%、賛成が約36.7%なので、参加率の推計値は約59.3%となる。実際の値よりも約7.5ポイント投票率を過小評価している。従来型は過大評価、リスト実験は過小評価なので「五十歩百歩」かもしれないが、実際の値と推定値のズレは従来型の方が大きい。実際の値に近いのはリスト実験の推定結果である。⁽¹⁷⁾

(17) 土屋・平井・小野(2007)の実験結果は、リスト項目が少ないとやや過大評価する値を、逆に項目数が多いと過小評価される推定となることを

表2 リスト実験による推定結果

	リスト実験			従来型			
	統制群	反対	賛成	参加比率		賛否比率	
0	23.3	20.1	17.6	参加	81.4	賛成	59.3
1	27.1	25.5	23.6				
2	45.8	39.2	35.6	棄権	17.6	反対	35.9
3	3.2	13.0	21.2				
4	0.0	1.3	1.1				
5		1.0	0.9	DK/NA	1.0	その他/DK	4.8
平均値	1.3041	1.5300	1.6714				
(Linearized S.E.)	(0.0339)	(0.0457)	(0.0432)				
推定%		22.59	36.73				
N	592	596	617	1805		1461	

注) すべてウェイト調整済みの値

同一サンプルに類似の質問を尋ねることとなるためキャリーオーバー効果が発生する可能性がある。しかし従来型の質問と比較して、リスト実験による参加率の推定値は約22.3ポイント小さい。つまりおよそ2割の回答者が「嘘をついていた」可能性があることを、この分析結果は示している。サンプリングバイアスとは異なるSDBが、投票参加について直接尋ねる際、たとえオンライン調査であっても生じる、ということである。

4.2 方向バイアスの検証

投票行動に係るSDBを、さらに一步踏み込んだ形で理解するには、投票方向との関係性を分析しなければならない。つまり「SDBは選好（投票先）表明と関係があるのか」という疑問にこたえる必要がある。以下では、都構想住民投票における賛否が、従来型とリスト実験とでどのように

示す。統制群のリストが3と4でこの相違が生じるとされているので、表1に示す結果は、本稿のリストが多いことによるものなのかもしれない。ただしこの点については本稿の議論の範囲を超えるため、これ以上の言及は避ける。

変化するかという観点から、SDBについて検討する。

図4は、従来型の質問による賛否比率の推定結果と、リスト実験による推定結果の差分を比較したものである。賛成、反対ともにリスト実験の方が、推定値は小さい。ただし、賛成の方が変動率は大きい。賛成の変動が10ポイント以上であるのに対して、反対のそれは約6ポイントである。両者の差は統計的に有意ではないが、やや反対の方が回答しづらい傾向にあるよう⁽¹⁸⁾だ。棄権者は反対よりも賛成と回答する可能性があることが、ここからはわかる。

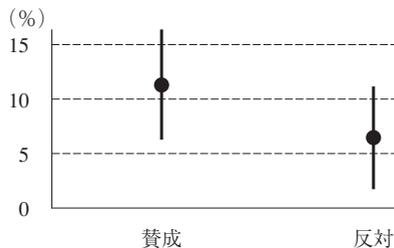


図4 賛否比率における従来型とリスト実験結果の推定値の差

注) ウェイト調整済みの値。縦棒は推定値の95%信頼区間

表3は、従来の直接尋ねる質問で「賛成」と回答した人と「反対」と回答した人を分けた上で、リスト実験による推定結果がどのように異なるのかを整理したものである。従来型の質問で賛成と回答した人の中で、賛成票を投じたと考えられるのは、その中の約74.4%である。他方、反対と回答した人の中で、反対票を投じた人は約77.3%である。比率でいうと、賛成も反対も、ともに7割から8割程度が「嘘をついていない」ことになる。一致率に賛否の差はほとんどない。「投票者」は正直に賛否の意思を表明

(18) 詳細は紙幅の都合上割愛するが、実際に賛成と反対のどちらが意見表明しづらいかを尋ねたところ、反対意見の方が表明しにくいと考える人が多いという結果が得られた。リスト実験の結果と整合的だといえる。

表3 一致率の推定結果

		リスト平均項目数			一致率
		統制群	反対	賛成	
従来型	賛成 (比率)	1.464 (61.5)	1.433 (56.9)	2.208 (59.6)	74.4%
	反対 (比率)	1.314 (34.2)	2.086 (37.9)	1.386 (35.6)	
	N	479	489	493	

注) すべてウェイト補正済みの値

(19) している。

参加だけではなく、どこに投票したのかという点まで加味して SDB について検討した結果、次の2点が明らかとなった。第1に、ある選好の表明が SDB の原因となる可能性がある。本稿の分析結果からいえば、SDB は、賛成回答を促す方向にかかりやすい。ただし両者の差は、統計的な有意差がでるほど大きなものではない。第2に、賛成と回答した人が反対へ、逆に反対と回答した人が賛成へとという傾向は確認できない。もちろん「参加するかどうか」という選好を表明する際、回答者は「嘘をつく」傾向にある。しかし投票に行った人は、投票先については、比較的正直に回答する。

4.3 性別と投票率の関係

前項までの検討を通じて、1) リスト実験による推定の方がより真値に近い推定結果となること、2) 参加したという SDB の背景には「民主的参加の規範」だけではなく「選好の表明の難易度」もある、という2点が

(19) 賛否が一致していない人のほとんどは、逆の回答、すなわち賛成なのに反対、あるいは反対なのに賛成票を投じているのではなく、棄権したということである。

明らかとなった。棄権者は、単に参加したと嘘をつくだけでなく、選好の表明が易しい方をあえて選び嘘をつく傾向にある。いずれにせよ、投票率の過大推計の背後にはSDBがあることを裏付ける結果が得られたと考えてよい。

投票参加率をより妥当性の高い形で把握することは、いうまでもなく適切な因果推論を行うことに繋がる。我々は社会に存在する因果関係を、様々なデータを用いて明らかにすることを試みる。しかし得られるデータが誤ったものであれば結果も誤ったものとなる。投票参加についても同様に、データ上の値が誤っているものなら、その規定要因や帰結に関する因果推論も誤ったものとなりやすい。

しかし具体的にどのような問題が生じるかは定かではない。そこで、この点を明らかにすべく、性別が投票参加とどのような関係にあるのかを分析する。性別に限定する理由は、実際の値が選挙管理委員会によって公表されている点に拠る。大阪市が公表している結果に従来型とリスト実験のどちらが近いのかを分析することで、どのようなミスが生じるのか、さらにリスト実験は妥当な方法なのかといった点を判断することができる。

図5は、性別と投票参加率の関係を整理したものである。男性の実際の投票率は約64.5%、女性は約69.0%である。女性の方が男性より投票に参加していることがわかる。しかし従来型の質問を用いると実態とは逆の結果になる。つまり男性の方が女性よりも投票参加率が高くなる。これに対してリスト実験による推定結果は、男性より女性の方が参加率が高い点で実態と整合的である。性別ごとの参加率推定値についても、より妥当な結果だといえる。男性については過小評価された値となっているが、女性については真値に近い推定値である。性別との関係については、従来型の質問を用いた場合ミスをする可能性が高い。これに対してリスト実験は実態に適合的な推論を可能とする。

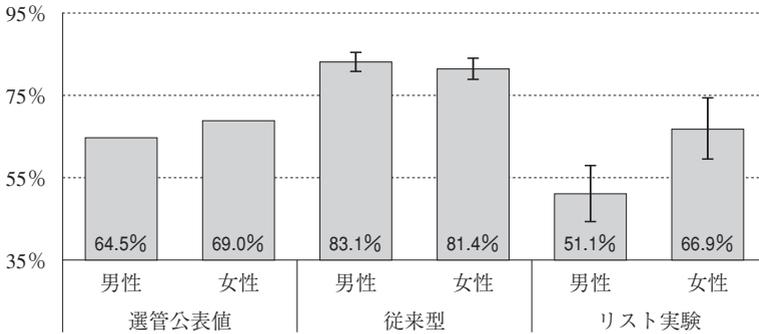


図5 性別と投票率の関係

注) すべてウェイト補正済みの値。縦棒は推定割合の95%信頼区間

大阪市選挙管理委員会は、確率抽出標本に基づき、年齢と投票率の関係についても真値に近い結果を公表している⁽²⁰⁾。そこで補足的な分析として、年齢と投票参加の関係を分析した結果についても、その概略を簡単に述べておく。

分析の結果を先取りして述べれば、従来型との比較という意味ではリスト実験による推定に軍配が上がるが、同時に、必ずしもリスト実験は万能ではないことを示す結果ともなった⁽²¹⁾。具体的にはリスト実験だと、若年層の投票率が極端に過小評価されることが明らかとなった。

詳しく述べると、まず、20-30代の投票率の選管公表値は約54.1%である。従来型は約70.3%，リストは約36.9%となり、どちらも不正確だった。従来型はおよそ16.2ポイント，リスト実験は33.4ポイント実態から乖離し

(20) 都構想住民投票の年齢別投票率についても、大阪市選挙管理委員会 HP 内の選挙資料として (<http://www.city.osaka.lg.jp/senkyo/cmsfiles/contents/0000321/321023/shukei.pdf>) 公開されている。ただし悉皆調査ではないため、結果には誤差が存在する。

(21) 年齢を細かく分けると誤差が大きくなってしまうため、20-30代、40-50代、60代以上とした。選管公表値もこれにあわせて再計算した。

ており、20-30代の投票率については、リスト実験による推定は妥当性を欠く。40-50代については、リスト実験による推定結果の方が「真値」に近い。従来型は約83.3%と10ポイント以上乖離した値となっているが、リスト実験は約73.8%である。60代以上についても、ややリスト実験による推定の方が精度が高い。20-30代と同じく、リスト実験による推定結果は過小評価されたものとなっているが、従来型の質問による推定結果が約92.3%と19ポイント程度乖離しているのに対して、リスト実験は乖離の程度が13.5ポイントであった。このように総体として見るとリスト実験による推定の方が、従来型よりも妥当性が高いといえる。しかし、若年層の傾向については注意する必要がある。

5. 結論

本稿では、投票参加の過大推計問題の実態を明らかにすると同時に、それを軽減する方法として考案されたリスト実験の有効性について検証した。加えて、リスト実験によってより適切な推論を行う意味を、性別と投票参加の関係という観点から明らかにした。

本稿の具体的な知見は、次の2点に集約される。第1は、民主的規範という側面だけではなく、意見表明のしやすさもSDBの原因となるということである。「棄権したと回答しにくい」ということに加えて、回答者はより「回答しやすい」選択肢を選ぶ傾向にある。第2はリスト実験が推定結果の改善に役立つ点である。特に性別との関係については、従来型の質問を用いた推定は明らかに妥当性を欠くものとなっていた。これに対してリスト実験による推定結果は実態に近いものであった。

また調査モードの変更だけではなく、さらなる方法論的工夫の必要性を示した点も重要である。従来の直接尋ねる形式だと調査員が介在しない場合であってもSDBが発生する。これは郵送調査などオンライン以外の調

査方法でも共通する問題である。事実、筆者はかつて大阪ダブル選における投票行動を郵送調査により実施したが（善教・石橋・坂本 2012）、推計された投票率は実際の値より高かった。

以上に見たようにリスト実験は、SDBを軽減させる方法として有効である。因果的推論の観点からも、リスト実験を用いることの意義は大きい。従来型の質問を用いる場合よりも、分析結果はより妥当性を増すものとなったからである。もっとも、本稿は先行研究の知見のすべてを否定するものではない。リスト実験を用いて投票方向まで加味した分析を行うのは困難である。たとえ不十分であっても、直接尋ねる形式を選択せざるを得ない場合はある。本稿はあくまで推定結果を鵜呑みにするのではなく、様々な観点から批判的に検討し直すことが必要であることを述べているに過ぎない。リスト実験は結果を補正する「補助的道具」として位置づけるべきである。しかし補助的道具であっても、それを用いる意義はある。

最後に、本稿の課題について数点述べておきたい。第1に本稿のデータは母集団から対象者を無作為に抽出したものではない。オンライン意識調査特有のサンプリングバイアスが存在している可能性は否定できない。第2に扱うテーマなどによって、どのような方法がSDBを軽減するのに役立つかは異なる。本稿ではこの点には踏み込んで検討できてない。RRTなど様々な方法との比較を行う必要がある。

付記：

本稿は文部科学省科学研究費補助金若手研究B（課題番号15K16995）による研究成果の一部である。

謝辞：

本稿は2015年度日本公共政策学会関西支部大会にて報告した内容に加筆・修正を施したものであり、報告の際にはフロアの先生方から、貴重なご指摘を頂いた。また、坂本治也氏（関西大学）と山野瞳氏（関西学院大学院生）

からも有益なご指摘を頂いている。ここに記して感謝申し上げる次第である。なお、残された誤りはすべて筆者の責にあることは言うまでもない。

参考文献

- 遠藤晶久（2012）「後援会動員と日本の有権者：世論調査モード間比較」『早稲田政治公法研究』第100号，pp. 1-14。
- 善教将大・石橋章市朗・坂本治也（2012）「大阪ダブル選挙の分析：有権者の選択と大阪維新の会支持基盤の解明」『関西大学法学論集』62巻3号，pp. 1019-1116。
- 田中愛治・日野愛郎編（2013）『世論調査の新しい地平：CASI方式世論調査』勁草書房。
- 土屋隆裕・平井洋子・小野滋（2007）「個別面接聴取法における Item Count法の諸問題と実用化可能性」『統計数理』第55巻，第1号，pp. 159-175。
- 西澤由隆・栗山浩一（2010）「面接調査における Social Desirability Bias：その軽減への full-scale CASI の試み」『レヴァイアサン』46号，pp. 51-74。
- 松林哲也（2015）「世論調査の回答率と投票率の推定誤差」『レヴァイアサン』第57号，pp. 96-117。
- 三浦麻子・小林哲郎（2015）「オンライン調査モニタの Satisfice に関する実験的研究」『社会心理学研究』31巻1号，pp. 1-12。
- 山口洋（2005）「キャリアオーバー効果をめぐる二つのジレンマ」『社会学部論集』第41号，pp. 49-60。
- 山田真裕（2009）「衆議院選挙制度改革の評価と有権者」『年報政治学』2009-I号，pp. 62-78。
- Blair, Graeme and Kosuke Imai (2012) “Statistical Analysis of List Experiments.” *Political Analysis*, Vol. 20, No. 1, pp. 47-77.
- Coutts, Elisabeth and Ben Jann (2011) “Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT).” *Sociological Methods Research*, Vol. 40, No. 1, pp. 169-193.
- Dunning, Thad (2012) *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press.
- Goertz, Gary and James Mahoney (2012) *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton: Princeton University Press (= 西川賢・今井真士 訳 [2015] 『社会科学のパラダイム論争：2つの文化の物語』勁草書房).

- Gonzalez - Ocantos, Ezequiel et al. (2012) "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science*, Vol. 56, Issue 1, pp. 202-217.
- Holbrook, Allyson L. and Jon A. Krosnick (2010a) "Social Desirability Bias in Voter Turnout Reports: Tests using the Item Count Technique." *Public Opinion Quarterly*, Vol. 74, Issue 1, pp. 37-67.
- (2010b) "Measuring Voter Turnout By Using The Randomized Response Technique: Evidence Calling Into Question The Method's Validity." *Public Opinion Quarterly*, Vol. 74, Issue 2, pp. 328-343.
- Imai, Kosuke, Bethany Park and Kenneth Greene (2015) "Using the Predicted Responses from List Experiments as Explanatory Variables in Regression Models." *Political Analysis*, Vol. 23, No. 2, pp. 180-196.
- Krumpal, Ivar et al. (2014) "Asking Sensitive Questions: A Critical Account of the Randomized Response Technique and Related Methods." Uwe Engel, Ben Jann, Peter Lynn, Annette Acherpenzeel and Patrick Sturgis eds. *Improving Survey Method: Lessons from Recent Research*. London: Routledge, pp. 122-136.
- Ottati, V. C. et al. (1989) "Cognitive and Affective Bases of Opinion Survey Responses." *Journal of Personality and Social Psychology*, Vol. 57, No. 3, pp. 404-415.
- Rosenfeld, Bryn, Kosuke Imai and Jacob Shapiro (forthcoming) "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science* (Online ver. <http://imai.princeton.edu/research/files/validate.pdf> 2015.8.26)
- Persson, Mikael J. and Maria Solevid (2014) "Measuring Political Participation: Testing Social Desirability Bias in a Web-Survey Experiment." *International Journal of Public Opinion Research*, Vol. 26, Issue 1, pp. 98-112.
- Selb, Peter and Simon Munzert (2013) "Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys." *Electoral Studies*, Vol. 32, Issue 1, pp. 186-196.
- Wolter, Felix and Bastian Laier (2014) "The Effectiveness of the Item Count Technique in Eliciting Valid Answers to Sensitive Questions. An Evaluation in the Context of Self-Reported Delinquency." *Survey Research Methods*, Vol. 8, No. 3, pp. 153-168.

補遺

1 無作為配分の妥当性

	統制群	処置群 1	処置群 2
男性	48.6%	47.7%	47.7%
女性	51.4%	52.3%	52.3%
$\chi^2=0.1479$	sig. (design-based) = 0.932		
20代	16.7%	13.5%	15.4%
30代	19.5%	17.5%	18.9%
40代	16.5%	17.6%	14.9%
50代	13.8%	14.6%	13.1%
60代以上	33.5%	36.9%	37.7%
$\chi^2=6.2292$	sig. (design-based) = 0.6487		
高卒以下	32.5%	28.2%	34.5%
専門・短大	23.6%	23.1%	24.5%
4年生大学	43.9%	48.7%	41.0%
$\chi^2=8.2477$	sig. (design-based) = 0.0734		
持家一戸建て	27.3%	30.9%	30.2%
その他	72.7%	69.1%	69.8%
$\chi^2=2.0358$	sig. (design-based) = 0.3903		
200万未満	10.3%	12.9%	12.5%
200-400万未満	28.3%	24.0%	24.2%
400-600万未満	21.4%	20.1%	23.2%
600-800万未満	13.4%	13.5%	11.4%
800-1000万未満	5.9%	8.8%	9.1%
1000万以上	9.6%	8.6%	6.8%
NA	11.0%	12.1%	12.9%
$\chi^2=15.7793$	sig. (design-based) = 0.2032		
関心：そう思う	21.5%	20.5%	19.0%
関心：ややそう思う	41.8%	41.4%	41.7%
関心：どちらとも言えない	19.2%	22.8%	21.8%
関心：あまりそう思わない	10.3%	8.2%	10.0%
関心：そう思わない	6.4%	6.2%	7.2%
DK	0.9%	0.8%	0.4%
$\chi^2=6.3372$	sig. (design-based) = 0.8244		

論
説

2 リスト項目平均値の信頼区間 (95, 99, 99.9% C. I.)

	95%下限	95%上限	99%下限	99%上限	99.9%下限	99.9%上限
統制群	1.237	1.371	1.215	1.393	1.190	1.418
処置群 1 (反対)	1.440	1.620	1.411	1.649	1.377	1.683
処置群 2 (賛成)	1.585	1.757	1.558	1.785	1.526	1.817

* 統制群と処置群 1 の差：1%水準，統制群と処置群 2 の差：0.1%水準で統計的に有意

Social Desirability Bias and Voting Turnout:
A Trial to Reduce Social Desirability Bias
Using List Experiment in Japan

Masahiro ZENKYO

One of the problems of social surveys on political behavior is the overreporting of voting behavior. The rates of survey reported turnout are almost lower than the estimated rates in the surveys. While this problem, to a certain extent, is caused by sampling bias, a recent studies have revealed that Social Desirability Bias (SDB)—the tendency of respondents to answer questions in a manner that will be favored by others—also matters. Many survey experiments such as Randomized Response Techniques have been devised, and many studies have tested the validity of these experimental designs. However, Japanese political scientists rarely conduct survey experiments to reduce SDB. Therefore, using a case of Japanese local election, this study examines the validity of “list experiment” to reduce SDB in voting turnout report. The findings of this study are as follows: (1) the difference in true turnout rates while using the list experiment is smaller than those while using conventional direct questions; (2) the result of analysis about the relation between gender and voting turnout by list experiment is more appropriate than those under direct questions. Although voting rates of women are higher than those of men, its inverse is shown in the result using direct questions.